GROWTH CURVE PREDICTION

by

Jack C. Lee and Seymour Geisser*

Technical Report No. 167

January 1972

University of Minnesota
Minneapolis, Minnesota

*This work was supported in part by an NIH Grant GM 14031.

0. Introduction.

In this paper we continue the study of the problem of prediction from the growth curve model initiated by Geisser (1970). A capsule history of the development of this model is given in the first paragraph of the aforementioned paper with appropriate references.

The model is

$$E(Y_{pxN}) = X_{pxN} \tau_{mxr} A_{rxN}$$

where τ is unknown, X and A are known matrices of ranks $m < p$ and $r < N$ respectively. Further, the columns of Y are independent and p-dimensional multinormal variates having a common unknown covariance matrix Σ.

Geisser (1970) considered this model from a Bayesian viewpoint and discussed posterior inference for τ and predictive inference for V, a future $p \times K$ observation matrix to be drawn from this model, i.e. K future p-dimensional vectors. This was examined under the assumption that Σ is an arbitrary unknown positive definite (p.d.) matrix and when $\Sigma = X\Gamma X' + Z\theta Z'$, which we term simple structure (S.S.), where $Z_{p \times p-m}$ is of rank $p-m$ such that $X'Z = 0$ and Γ, θ are arbitrary unknown p.d. matrices.

The main burden of this paper is to examine the problem of partial or conditional predictions from this model. In the previous paper prediction was restricted to V, the whole set of future observations. If we let $V = \begin{pmatrix} V^{(1)} \\ V^{(2)} \end{pmatrix}$ where $V^{(i)}$ is $p_i \times K$, $i = 1,2$ $p_1 + p_2 = p$, then our interest is in predicting $V^{(2)}$ after observing $V^{(1)}$, and of course Y. A typical problem of this kind may be conceived as follows;

A particular attribute is measured on a random sample of N children from some specified population on their 8th, 9th, 10th and 11th birthdays. We wish to predict this attribute of a child from this population for his 11th birthday when he has just turned 10 using his measurements on his previous three birthdays and the measurements of the other N children on their 8th through 11th birthdays. One could also use this method to "predict" for another child his attribute at age 9, assuming the measurement was lost or for some reason not taken, based on values at the other three ages. This latter case is generally called the "missing value problem" but with regard to our model it is still conditional prediction and is handled in exactly the same manner as the previous problem. It is clear of course that this includes the case of predicting the observation matrix V by setting $p_1 = 0$.

The $E(V)$ in the predictive distribution of V can be used for the estimation of V, and it is easily obtained. However an explicit analytic evaluation of $E\left(V^{(2)} \middle| V^{(1)}\right)$ seems to be intractable. We shall however present approximations and for the case $K = 1$ numerical procedures for calculating this value when the covariance is of simple structure. We also delineate certain low-dimensional search proceedures which are capable of calculating the mode of $V^{(2)}$ in the conditional distribution of $V^{(2)}$ given $V^{(1)}$ when $K = 1$. The approximations we shall consider depend on estimates of Σ or the covariance matrix of the predictive distribution of V. We shall deal with these matters pertaining to the arbitrary covariance case in section 1. In section 2 we shall discuss approximations and in the third section develop the modal calculations for the aforementioned case. Section 4 introduces the simple structure case and presents a test of this case versus the arbitrary case. Sections 5-8 then deal with predictions when simple structure obtains.

1. Estimates of Σ and the Predictive Covariance.

For the sake of convenience we shall deal with the pseudo-augmented model

$$(1.1)\qquad E(Y) = (X,Z)\binom{\tau}{0}A$$

The likelihood function of τ and Σ is

$$(1.2)\qquad L(\tau,\Sigma) = (2\pi)^{-\frac{pN}{2}} |\Sigma|^{-\frac{N}{2}} \exp\left\{ -\tfrac{1}{2}\mathrm{tr}\left[\binom{B}{Z'}\Sigma(B',Z)\right]^{-1} \begin{bmatrix}BY-\tau A\\ Z'Y\end{bmatrix} \begin{bmatrix}BY-\tau A\\ Z'Y\end{bmatrix}'\right\}$$

where

$$(1.3)\qquad B = (X'X)^{-1} X'$$

First we shall need the following lemma:

<u>Lemma 1.1</u> (Rao 1965, Khatri 1966)

Let $D_{p\times q}$ and $C_{p\times p-q}$ be of ranks q and p-q respectively such that $C'D = 0$. If S is p.d., then

$$S^{-1} - S^{-1}D\,(D'S^{-1}D)^{-1}\,D'S^{-1} = C\,(C'SC)^{-1}\,C' .$$

Now following more or less along the lines of Khatri (1966) we let

$$(1.4)\qquad \Lambda^{-1} = \left[\binom{B}{Z'}\Sigma(B',Z)\right]^{-1} = \begin{pmatrix}\Lambda^{11} & -\Lambda^{11}\eta\\ \eta'\Lambda^{11} & \eta'\Lambda^{11}\eta+(Z'\Sigma Z)^{-1}\end{pmatrix}$$

where

$$(1.5)\qquad \begin{cases}\Lambda^{11} = (B\Sigma B' - \eta Z'\Sigma B)^{-1}\\ \eta = B\Sigma Z(Z'\Sigma Z)^{-1}\end{cases}.$$

We can reparametrize the likelihood function as

$$(1.6)\qquad L(\tau,\eta, X'\Sigma^{-1}X, Z'\Sigma Z) = (2\pi)^{-\frac{pN}{2}} |Z'\Sigma Z|^{-\frac{N}{2}} |X'\Sigma^{-1}X|^{\frac{N}{2}} |(B',Z)|^{N}$$
$$\cdot\exp[-\tfrac{1}{2}\mathrm{tr}\, X'\Sigma^{-1}X\{[(\tau,\eta)-BYQ^{*\prime}(Q^*Q^{*\prime})^{-1}]Q^*Q^{*\prime}[(\tau,\eta)-BYQ^{*\prime}(Q^*Q^{*\prime})^{-1}]'$$
$$+BY[I-Q^{*\prime}(Q^*Q^{*\prime})^{-1}Q^*]\,Y'B'\}]$$
$$\cdot\exp[-\tfrac{1}{2}\mathrm{tr}\,(Z'\Sigma Z)^{-1}\,Z'YY'Z]$$

where

$$Q^* = \begin{pmatrix} A \\ Z'Y \end{pmatrix}$$

By the utilization of Lemma 1.1 it can be shown that the MLE's are

$$(1.7) \quad \begin{cases} \hat{\tau} = (X'S^{-1}X)^{-1} X'S^{-1} YA'(AA')^{-1} \\ \widehat{(X' \Sigma^{-1}X)}^{-1} = N^{-1} BY[I-Q^{*\prime}(Q^*Q^{*\prime})^{-1} Q^*] Y'B' \\ \widehat{(Z'\Sigma Z)} = N^{-1}Z'YY'Z \\ \hat{\eta} = BSZ (Z'SZ)^{-1} \end{cases}$$

where

$$(1.8) \qquad S = Y[I-A'(AA')^{-1}A] \quad Y'$$

From (1.7) and Lemma 1.1 we have the MLE of Λ^{-1} , with $D = (Z'Z)^{-1} Z'$,

$$(1.9) \qquad \hat{\Lambda}^{-1} = N\begin{pmatrix} X' \\ D \end{pmatrix} S^{-1} (X,D') + N \begin{bmatrix} 0 & 0 \\ 0 & (Z'YY'Z)^{-1} - (Z'SZ)^{-1} \end{bmatrix}$$

which implies that the MLE of Σ^{-1} is

$$(1.10) \qquad \hat{\Sigma}^{-1} = NS^{-1} + NZ[(Z'YY'Z)^{-1} - (Z'SZ)^{-1}] Z' .$$

We thus see that the MLE of Σ is

$$(1.11) \qquad \hat{\Sigma} = N^{-1}\{S^{-1} + Z[(Z'YY'Z)^{-1} - (Z'SZ)^{-1}] Z'\}^{-1}$$
$$= N^{-1}S \left(I-SZ[(Z'SZ)^{-1} - (Z'YY'Z)^{-1}] Z'\right)^{-1}$$

which indicates how $\hat{\Sigma}$ is a function of the simple unbiased estimate $(N-r)^{-1}S$ of Σ .

Alternatively, the MLE of Σ is, of course,

$$(1.12) \qquad \hat{\Sigma} = N^{-1} (Y-X\hat{\tau}A)(Y-X\hat{\tau}A)' .$$

A preference amongst the various forms above will depend on the computational proceedure.

We now consider the posterior expectation of Σ . The joint posterior density of τ and Σ^{-1} is, with a convenient prior density

$$g(\tau,\Sigma^{-1}) \propto |\Sigma|^{(p+1)/2} \quad , \quad \text{Geisser and Cornfield (1963)} ,$$

$$(1.13) \qquad P(\tau,\Sigma^{-1}) \propto |\Sigma^{-1}|^{(N-p-1)/2} \exp[-\tfrac{1}{2}\mathrm{tr}\ \Sigma^{-1}(Y-X\tau A)(Y-X\tau A)']$$

From (1.13) Geisser (1970) showed that a posteriori

$$(1.14) \qquad P(\tau) \propto |(X'S^{-1}X)^{-1} + (\tau-\hat{\tau})G(\tau-\hat{\tau})'|^{-\frac{N}{2}}$$

where

$$(1.15) \qquad \begin{cases} G^{-1} = (AA')^{-1} + T_2'(Z'SZ)^{-1}T_2 \\ T_2 = Z'YA'(AA')^{-1} \end{cases}$$

The posterior density of τ as given in (1.14) is the general determinantal density and will be denoted by $D(\cdot;\hat{\tau}, G, (X'S^{-1}X)^{-1}, N)$. We shall say that B is distributed as $D(\cdot;\Delta,\Lambda,\Sigma,N)$ if

$$(1.16) \qquad f(B) = \frac{C_{m,\nu}\, \pi^{-rm/2}|\Sigma|^{\nu/2}\, |\Lambda|^{m/2}}{C_{m,N}\, |\Sigma + (B-\Delta)\Lambda(B-\Delta)'|^{N/2}}$$

where Σ is $m \times m$ and p.d., Λ is $r \times r$ and p.d., and B and Δ are $m \times r$ and $\nu = N-r \geq m \geq 1$, and

$$(1.17) \qquad C^{-1}_{m,\nu} = \pi^{m(m-1)/4}\prod_{i=1}^{m} \Gamma\left(\frac{\nu+1-i}{2}\right) ,$$

Geisser (1966). We note in passing that the general determinantal density is also referred to as the matrix T density, Dickey (1967), and the multivariate T density is just a special case. For convenience and subsequent reference, a multivariate T distribution, denoted by $T(\cdot;\mu,\Sigma,N)$, whose density is defined as

$$f(B) = (\pi\nu)^{-\frac{p}{2}}\ \Gamma^{-1}\left(\frac{\nu}{2}\right)\Gamma\left(\frac{N}{2}\right)|\Sigma|^{-\frac{1}{2}}[1 + (B-\mu)'(\nu\Sigma)^{-1}(B-\mu)]^{-\frac{N}{2}}$$

where B is $p \times 1$, $\nu = N - p$.

In order to compute the posterior expectation of Σ , we shall need a few preliminary results.

<u>Lemma 1.2</u> (Lachenbruch 1968)

If $S_{p \times p} \sim W(\cdot;\Sigma,\nu)$, then

$$E(S^{-1}) = (\nu-p-1)^{-1}\,\Sigma^{-1}$$

where $W(\cdot;\Sigma,\nu)$ stands for the Wishart distribution with parameters Σ and ν.

<u>Theorem 1.1</u>.

If V and Δ are $p \times K$ such that

$$f(V) \propto |\Sigma + (V-\Delta)\Lambda(V-\Delta)'|^{-\frac{N}{2}}, \text{ i.e,}$$

(1.18) $$V \sim D(\cdot;\Delta,\Lambda,\Sigma,N),$$

then

$$E(V) = \Delta, \text{ and}$$

$$Cov(V) = (N-p-K-1)^{-1}\,\Sigma \otimes \Lambda^{-1},$$

if they exist, where $\otimes$ denotes the Kronecker product and $Cov(V)$ stands for the variance-covariance matrix among the rows of V, as defined by Anderson (1958 p. 182).

<u>Proof</u>:

The expectation is trivial while the covariance is obtained from the well-known identity

(1.19) $$Cov(V) = E_S\,[Cov(V|S)] + Cov_S(E[V|S]),$$

the application of Lemma 1.2 and the fact that (1.18) can be represented as

(1.20) $$\begin{cases} F(V|S) = N(\cdot;\Delta,S^{-1} \otimes \Lambda^{-1}) \\ S \sim W(\cdot;\Sigma^{-1}, N-\ ) \end{cases}$$

where $F(\cdot|\cdot)$ denotes the distribution of function. We note that the posterior expectation of τ, if it exists, is $\hat{\tau}$ as given in (1.7).

<u>Theorem 1.2.</u>

If $V_{p \times K} \sim D(\cdot;\Delta,\Lambda,\Sigma,N)$, then

$$V' \sim D(\cdot;\Delta',\Sigma^{-1},\Lambda^{-1},N).$$

<u>Proof:</u>

This follows from the representation (1.20) and the fact that

$$X \sim N(\cdot;\mu,\Sigma_1 \otimes \Sigma_2) \text{ implies}$$
$$X' \sim N(\cdot;\mu',\Sigma_2 \otimes \Sigma_1) .$$

Theorem 1.3.

Let C_1 be $K \times \ell$ of rank $\ell \le K$, C_2 be $s \times p$ of rank $s \le p$, A be any p.d. constant matrix, and V be distributed as (1.18). Then

(i) $VC_1 \sim D(\cdot;\Delta C_1, (C_1'\Lambda^{-1}C_1)^{-1}, \Sigma, N + \ell - k)$

(ii) $C_2V \sim (D^{\cdot};C_2\Delta, \Lambda, C_2\Sigma C_2', N + s-p)$

(iii) $E[VAV'] = (N - p - K - 1)^{-1} (\mathrm{tr}\ \Lambda^{-1}A) \Sigma + \Delta A \Delta'$

Proof:

(i) follows from the representation (1.20)

(ii) follows from Theorem 1.2 and (i)

(iii) follows from Theorems 1.1, 1.2 and (i) .

We are now in a position to find the posterior expectation of Σ . From (1.13) and (1.14) we have

$$\begin{cases} F(\Sigma^{-1}|\tau) = W(\cdot;[(Y - X\tau A)(Y - X\tau A)']^{-1}, N) \\ \tau \sim D(\cdot;\hat{\tau}, G, (X'S^{-1}X)^{-1}, N) . \end{cases} \tag{1.21}$$

Thus by Lemma 1.2 and Theorem 1.3 (iii),

$$\begin{aligned} E(\Sigma) = (N - p - 1)^{-1} \{&(Y - X\hat{\tau}A)(Y - X\hat{\tau}A)' \\ &+ (N-m-r-1)^{-1} X(X'S^{-1}X)^{-1} X'[\mathrm{tr}\ G^{-1} AA']\} . \end{aligned} \tag{1.22}$$

Comparison of (1.22) with MLE of Σ as given in (1.12) we see that the difference $\Delta = E(\Sigma) - \hat{\Sigma}$ is always positive definite.

We next consider the covariance matrix of the predictive distribution of V . Since V is assumed to be drawn from the growth curve model we have

$$F(V|\tau, \Sigma^{-1}) = N(\cdot;X\tau F, \Sigma \otimes I_K) \tag{1.23}$$

where

$F_{r \times K}$ is a known design matrix, usually formed by some columns of A. From (1.19), (1.22) and Theorem 1.1 we have

$$(1.24) \qquad Cov(V) = (N-p-1)^{-1} (Y-X\hat{\tau}F)(Y-X\hat{\tau}F)' \otimes I_K$$
$$+ (N-r-m-1)^{-1}X(X'S^{-1}X)^{-1}X' \otimes [(N-p-1)^{-1}(tr\ G^{-1}AA')I_K + F'G^{-1}F].$$

which is free of Z and has been obtained without knowing the predictive density of V . These results shall be useful for the approximations developed in the next section.

2. Approximations.

As noted by Geisser (1970) that it may be extremely troublesome to obtain a $1 - \alpha$ probability region for a future observation matrix V which is of minimum hypervolume. If V is partitioned as $V = \begin{pmatrix} V^{(1)} \\ V^{(2)} \end{pmatrix}$ and we are interested in a marginal predictive region for $V^{(i)}$, $i = 1,2$ or a conditional predictive region for $V^{(2)}|V^{(1)}$, then the previously mentioned difficulty is increased.

In this section we will present some approximations to the predictive density of V . Since the expectation and the covariance matrix of V has been obtained, we can approximate the predictive distribution by

$$(2.1) \qquad F(V_{p \times 1}) \simeq N(\cdot;\ X\hat{\tau}F,\ \Sigma_V)$$

where Σ_V is the covariance matrix of the predictive distribution as given in (1.24), since V tends to normality as N increases.

From the above approximation we can obtain marginal predictions for $V^{(i)}$ as well as conditional predictions for $V^{(2)}$ given $V^{(1)}$ using the standard normal theory.

In the rest of this section we will assume that Σ is essentially known, i.e. replaced by an estimate; unbiased estimate, MLE or the posterior expectation.

The density of V conditional on τ and Σ^{-1} is

(2.2) $$f(V|\tau,\Sigma^{-1}) \propto |\Sigma^{-1}|^{K/2} \exp[-\tfrac{1}{2}\mathrm{tr}\ \Sigma^{-1} (V - X\tau F)(V - X\tau F)'] \ .$$

with Σ known the posterior density of τ given Σ^{-1} is

(2.3) $$P(\tau|\Sigma^{-1}) \propto |\Sigma^{-1}|^{(N-p-1)/2} \exp[-\tfrac{1}{2}\mathrm{tr}\ \Sigma^{-1} (Y - X\tau A)(Y - X\tau A)'] \ .$$

Combining (2.2) with (2.3) and making transformation $\Sigma^{-1} = (B',Z)\ \Lambda^{-1}(B',Z)'$ and then applying Lemma 1.1 we obtain

(2.4) $$f(V,\tau|\Lambda^{-1}) \propto |\Lambda^{-1}|^{(N+K-p-1)/2}$$
$$\cdot\exp\{-\tfrac{1}{2}\mathrm{tr}\ \Lambda^{-1}[W - \binom{\tau}{0}H][W - \binom{\tau}{0}H]'\}$$

where

(2.5) $$\begin{cases} W = \begin{pmatrix} B \\ Z' \end{pmatrix} (Y,V) \\ H = (A,F) \end{cases}$$

and Λ^{-1} was given in (1.4).

Utilization of the identity (4.12) of Geisser (1970) and Lemma 1.1 yields, after same algebra and integration w.r.t. τ ,

(2.6) $$f(V|\Sigma^{-1}) \propto \exp[-\tfrac{1}{2}\mathrm{tr}\ (Z'\Sigma Z)^{-1} V_2 V_2'$$
$$\cdot\exp -\tfrac{1}{2}\mathrm{tr}\ X'\Sigma^{-1} X(V_1-Q)\ M(V_1-Q)']$$

where

(2.7) $$\begin{cases} Q = B\hat{V} + \eta(V_2 - Z'\hat{V}) \\ \hat{V} = YA'(AA')^{-1}\ F \end{cases}$$

and η was defined in (1.5).

We now recognize that (2.7) can be written as

(2.8) $$f(V|\Sigma^{-1}) = f(V_1|V_2,\Sigma^{-1})\ f(V_2|\Sigma^{-1})$$

where

$$(2.9)\quad \begin{cases} F(V_1|V_2,\ \Sigma^{-1}) = N(\cdot;Q,(X'\Sigma^{-1}X)^{-1} \otimes M^{-1}) \\ F(V_2|\Sigma^{-1}) = N(\cdot;\ 0,\ Z'\Sigma Z\otimes I_K) \end{cases} .$$

Hence, we have

$$(2.10)\quad F(V|\Sigma^{-1}) = N(\cdot;\ \mu_a,\Sigma_a)$$

where

$$(2.11)\quad \begin{cases} \mu_a = X(X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1}\hat{V} \\ \Sigma_a = X(X'\Sigma^{-1}X)^{-1} X'\otimes M^{-1} + [XB\Sigma Z(Z'\Sigma Z)^{-1} Z'\Sigma B'X' + D'Z'\Sigma B'X' \\ \qquad + XB\Sigma ZD + D'Z'\Sigma ZD] \otimes I_K \\ M = I - F'(HH')^{-1} F \end{cases}$$

We thus see, with Σ replaced by its estimate, that inference on V as well as $V^{(2)}$ conditional on $V^{(1)}$ can be obtained through the standard normal theory, with appropriate rearrangement of V and the corresponding covariance matrix Σ_a .

3. Predictive Modes.

It has been noticed by Geisser (1970) that the expectation $X\hat{\tau}F$ of the predictive distribution of V is not the mode. In this section we will present a procedure for obtaining the predictive mode of V as well as the mode of the conditional distribution of $V^{(2)}$ given $V^{(1)}$ when $K = 1$.

The predictive density of V as given by Geisser (1970, eq. 4.21) can be written as

$$(3.1)\quad f(V) \propto [1 + (V - X\hat{\tau}F)'\ Q^{(1)}\ (V - X\hat{\tau}F)]^{-(N+1-m)/2}$$
$$\cdot[M^{-1} + (V - \hat{V})'\ Q^{(2)}(V - \hat{V}) + (V - X\hat{\tau}F)'Q^{(3)}(V - X\hat{\tau}F)]^{-(N+1-r)/2}$$
$$\cdot[M^{-1} + (V - \hat{V})'\ Q^{(2)}(V - \hat{V})]^{(N+1-r-m)/2}$$

where

$$(3.2)\qquad \begin{cases} Q^{(1)} = Z\ (Z'YY'Z)^{-1}\ Z' \\ Q^{(2)} = Z\ (Z'SZ)^{-1}\ Z' \\ Q^{(3)} = S^{-1}\ X\ (X'S^{-1}X)^{-1}\ X'S^{-1} \end{cases}$$

From (3.1) we can find the mode V_m of the predictive distribution by utilizing a suggestion of Lindley (1970), this is accomplished by

$$\min_V\ (V - X\hat{\tau}F)'\ Q^{(1)}\ (V - X\hat{\tau}F)$$

subject to

$$(3.3)\qquad \begin{cases} (V - \hat{V})'\ Q^{(2)}\ (V - V) = k_2 \\ (V - X\hat{\tau}F)'\ Q^{(3)}\ (V - X\hat{\tau}F) = k_3 \end{cases} \quad ;$$

or, equivalently,

$$(3.4)\qquad \min_V q = \min_V \{(V - X\hat{\tau}F)'Q^{(1)}\ (V - X\hat{\tau}F)$$
$$+ \lambda_2\ [\ (V - \hat{V})'Q^{(2)}\ (V - \hat{V}) - k_2]$$
$$+ \lambda_3\ [\ (V - X\hat{\tau}F)'Q^{(3)}\ (V - X\hat{\tau}F) - k_3]\}$$

If we set $\frac{\partial q}{\partial V} = 0$, this then implies

$$(3.5)\qquad V_m = X\hat{\tau}F - (Q^{(1)} + \lambda_2 Q^{(2)} + \lambda_3 Q^{(3)})^{-1} \left[\lambda_2 Q^{(2)} (X\hat{\tau}F - V)\right] .$$

Hence the modal value of the predictive distribution will satisfy (3.5). Thus the mode of this distribution can be obtained in a two-dimensional search over λ_2 and λ_3 , i.e. the mode of the predictive distribution has been reduced from a p-dimensional search to a two-dimensional one. We note in passing that $Q^{(1)} + Q^{(2)} + Q^{(3)} = Q^{(1)} + S^{-1}$ and hence is p.d.

In practice we can either substitute V_m into (3.1) and search for the V which maximizes f(V) or consider $f(V_m)$ as a function of λ_2 and λ_3 , i.e. $f(V_m) = h(\lambda_2,\lambda_3)$. Then setting $\frac{\partial}{\partial \lambda_i} \log h(\lambda_2,\lambda_3)=0$

for $i = 2,3$, we have two equations

$$(3.6) \qquad \begin{cases} \lambda_2 = h_1(\lambda_2,\lambda_3) \\ \lambda_3 = h_2(\lambda_2,\lambda_3) \end{cases} .$$

Thus numerically we can obtain λ_2^*, λ_3^* such that $h(\lambda_2,\lambda_3)$ is maximized and hence $V_m(\lambda_2^*,\lambda_3^*)$ is the mode.

Since the mode of the predictive distribution is obtainable, we can also approximate the distribution by

$$(3.7) \qquad F(V_{p \times 1}) \simeq N(\cdot;\ V_m,\ \Sigma_V^*)$$

where V_m is the mode of the predictive distribution as given in (3.5) and

$$(3.8) \qquad \Sigma_V^* = E[(V-V_m)(V-V_m)'] = Cov(V) + X\hat{\tau}FF'\hat{\tau}'X' - V_mF'\hat{\tau}'X' - X\hat{\tau}FV_m' + V_m V_m' .$$

We next consider conditional predictive inference on $V^{(2)}$ given $V^{(1)}$.

<u>Lemma 3.1.</u>

If $Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$, $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$, where A is symmetric positive semi-definite, then

$$(3.9) \qquad Y'AY = (Y_2 + A_{22}^- A_{21} Y_1)'\ A_{22}(Y_2 + A_{22}^- A_{21} Y_1) + Y_1'\ A_{11.2} Y_1$$

where

$$(3.10) \qquad A_{11.2} = A_{11} - A_{12}A_{22}^-A_{21} ,$$

and A_{22}^- is the weak generalized inverse of A_{22} satisfying

$$(3.11) \qquad \begin{cases} A_{22}\ A_{22}^-\ A_{22} = A_{22} \\ A_{22}^-\ A_{22}\ A_{22}^- = A_{22}^- \\ (A_{22}^-\ A_{22})' = A_{22}^-\ A_{22} \end{cases}$$

<u>Proof:</u>

It follows by the fact that $A_{22}\ A_{22}^-\ A_{21} = A_{21}$ as noted by Zelen and Federer (1965).

Now by applying Lemma 3.1 and the fact that $f(V^{(2)}|V^{(1)}) \propto f(V)$, we have from (3.1)

$$f(V^{(2)}|V^{(1)}) \propto [K_1 + (V^{(2)} - \mu_{2\cdot 1})' Q_{22}^{(1)} (V^{(2)} - \mu_{2\cdot 1})]^{-(N+1-m)/2} \tag{3.12}$$
$$\cdot [K_2 + (V^{(2)} - \tilde{\mu}_{2\cdot 1})' Q_{22}^{(2)} (V^{(2)} - \tilde{\mu}_{2\cdot 1}) + (V^{(2)} - \dot{\mu}_{2\cdot 1})' Q_{22}^{(3)} (V^{(2)} - \dot{\mu}_{2\cdot 1})]^{-(N+1-r)/2}$$
$$\cdot [K_3 + (V^{(2)} - \tilde{\mu}_{2\cdot 1})' Q_{22}^{(2)} V^{(2)} - \tilde{\mu}_{2\cdot 1})]^{(N+1-r-m)/2}$$

where

$$\left\{\begin{array}{l}
K_1 = 1 + (V^{(1)} - X^{(1)}\hat{\tau}F)' Q_{11\cdot 2}^{(1)} (V^{(1)} - X^{(1)}\hat{\tau}F) \\
\mu_{2\cdot 1} = X^{(2)}\hat{\tau}F - Q_{22}^{(1)-} Q_{21}^{(1)} (V^{(1)} - X^{(1)}\hat{\tau}F) \\
K_2 = M^{-1} + (V^{(1)} - \hat{V}^{(1)})' Q_{11\cdot 2}^{(2)} (V^{(1)} - \hat{V}^{(1)}) + (V^{(1)} - X^{(1)}\hat{\tau}F)' Q_{11\cdot 2}^{(3)} (V^{(1)} - X^{(1)}\hat{\tau}F) \\
\hat{V} = \begin{pmatrix} \hat{V}^{(1)} \\ \hat{V}^{(2)} \end{pmatrix} \\
X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} \\
\tilde{\mu}_{2\cdot 1} = \hat{V}^{(2)} - Q^{(2)-} Q_{21}^{(2)} (V^{(1)} - \hat{V}^{(1)}) \\
Z = \begin{pmatrix} Z^{(1)} \\ Z^{(2)} \end{pmatrix} \\
\dot{\mu}_{2\cdot 1} = X^{(2)}\hat{\tau}F - Q_{22}^{(3)-} Q_{21}^{(3)} (V^{(1)} - X^{(1)}\hat{\tau}F) \\
K_3 = M^{-1} + (V^{(1)} - \hat{V}^{(1)})' Q_{11\cdot 2}^{(2)} (V^{(1)} - \hat{V}^{(1)}) \\
Q^{(i)} = \begin{pmatrix} Q_{11}^{(i)} & Q_{12}^{(i)} \\ Q_{21}^{(i)} & Q_{22}^{(i)} \end{pmatrix} \\
Q_{11\cdot 2}^{(i)} = Q_{11}^{(i)} - Q_{12}^{(i)} Q_{22}^{(i)-} Q_{21}^{(i)} ,
\end{array}\right. \tag{3.13}$$

and $Q_{22}^{(i)-}$ is the weak generalized inverse of $Q_{22}^{(i)}$, for $i = 1,2,3$.

It is easily seen that (3.12) is exactly the same form as (3.1).

Hence similar to the previous arguments the modal value of the distribution of $V^{(2)}$ given $V^{(1)}$ will satisfy

(3.14) $$V_m^{(2)} = (Q_{22}^{(1)}+\ell_2 Q_{22}^{(2)}+\ell_3 Q_{22}^{(3)})^{-1} (Q_{22}^{(1)} \mu_{2\cdot 1}+\ell_2 Q_{22}^{(2)} \tilde{\mu}_{2\cdot 1}+\ell_3 Q_{22}^{(3)} \dot{\mu}_{2\cdot 1}) .$$

Thus the mode of conditional predictive distribution can be obtained in a two-dimensional search over ℓ_2 and ℓ_3 .

4. Test For Simple Versus Arbitrary Structure.

Rao (1967) showed that the necessary and sufficient condition that the least squares estimator of τ in the generalized growth curve model is the same as that for $\Sigma = \sigma^2 I$ is $\Sigma = X\Gamma X' + Z\theta Z' + \sigma^2 I$. Geisser (1970) noted that without loss of generality this simple structure covariance can be written as $\Sigma = X\Gamma X' + Z\theta Z'$ where Γ, θ are unspecified and X, Z are defined as before.

In this section we shall present a likelihood ratio test for

H_o: $\Sigma = X\Gamma X' + Z\theta Z'$, Γ and θ are p.d. and unspecified

vs.

H_1: Σ arbitrary p.d.

By virtue of the fact that under H_o , $\Sigma^{-1} = B'\Gamma^{-1}B + D'\theta^{-1}D$, and the identity $(BY - \tau A)(BY - \tau A)' = BSB' + (\tau - T_1)AA'(\tau - T_1)'$ where $T_1 = BYA'(AA')^{-1}$, the likelihood function of τ, Γ and θ is

(4.1) $$L(\tau,\Gamma,\theta) = (2\pi)^{-pN/2} \, |(B',D')|^N \, |\Gamma|^{-\frac{N}{2}} \, |\theta|^{-\frac{N}{2}}$$
$$\cdot \exp(-\tfrac{1}{2}\mathrm{tr}\, \Gamma^{-1} [BSB' + (\tau - T_1)\, AA'\, (\tau - T_1)'])$$
$$\cdot \exp(-\tfrac{1}{2}\mathrm{tr}\, \theta^{-1}\, DYY'D') .$$

From (4.1) we obtain the MLE's

(4.2) $$\hat{\tau} = T_1,\ \hat{\Gamma} = N^{-1}BSB',\ \hat{\theta} = N^{-1}DYY'D' .$$

Hence

(4.3) $$\max_{\tau,\Gamma,\theta} L(\tau,\Gamma,\theta) = (2\pi)^{-pN/2} \, |(B',D')|^{N} |N^{-1}BSB'|^{-N/2} |N^{-1}DYY'D'|^{-N/2} \cdot \exp\left[\frac{Np}{2}\right] .$$

From (4.2) the MLE of Σ under H_o is $\hat{\Sigma} = N^{-1}XBSB'X' + N^{-1}ZDYY'D'Z'$ which is incidentally a biased estimate of Σ. An unbiased estimate is easily supplied, if desirable, and is $(N-r)^{-1}XBSB'X' + N^{-1}ZDYY'D'Z'$.

From (1.7) we have under H_1

(4.4) $$\max L(\tau,\Sigma) = (2\pi)^{-pN/2} |(B',Z)|^{N} |N^{-1}Z'YY'Z|^{-N/2} |NX'S^{-1}X|^{N/2} \exp\left[\frac{-Np}{2}\right]$$

Combining (4.3) and (4.4) we obtain the likelihood ratio test statistic

(4.5) $$\tilde{\lambda} = \frac{|(X'S^{-1}X)^{-1}|}{|BSB'|} .$$

By Lemma 1.1 we have the identity $(X'S^{-1}X)^{-1} = BSB' - BSZ(Z'SZ)^{-1}Z'SB'$. Since $S \sim W(\cdot;\Sigma, N-r)$, we have

(4.6) $$\binom{B}{Z'} S\,(B',Z) \sim W(\cdot;\Sigma^*,N-r)$$

where

(4.7) $$\Sigma^* = \begin{pmatrix} B\Sigma B' & B\Sigma Z \\ Z'\Sigma B' & Z'\Sigma Z \end{pmatrix} = \begin{pmatrix} \Sigma^*_{11} & \Sigma^*_{12} \\ \Sigma^*_{21} & \Sigma^*_{22} \end{pmatrix}$$

We shall also need the following Theorem:

<u>Theorem 4.1.</u>

$\Sigma^*_{12} = 0$ if and only if $\Sigma = X\Gamma X' + Z\theta Z'$, where Γ,θ are argitrary p.d. matrices.

<u>Proof:</u>

This follows by the fact that $X'Z = 0$ and $X(X'X)^{-1}X' + Z(Z'Z)^{-1}Z' = I$.

From (4.6) and theorem 4.1 we see that the distribution of $\tilde{\lambda}$ under H_o is the same as that of a U statistic as given in Anderson (1958 p.243). Applying his result, we have under H_o

(4.8) $\qquad \tilde{\lambda} \sim U_{m,p-m,N-(p-m)-1}$.

We next present the asymptotic non-null distribution of $\tilde{\lambda}$.

Lemma 4.1 (Cramér (1946 p. 366); Olkin and Press (1969 p. 1364)) .

Suppose $S \sim W(\cdot;\Sigma,n)$. Let $\theta = n^{-1}\Sigma^{-\frac{1}{2}}S\Sigma^{-\frac{1}{2}}$ and $h(\theta) = g(n\Sigma^{\frac{1}{2}}\theta\Sigma^{\frac{1}{2}})$. If h is scalar invariant and satisfies condition (1) in Cramér (1946 p. 353), then $\mathcal{L}\{\sqrt{n}[h(\theta) - h(0)]\} \to N(\cdot;0, 2\mathrm{tr}\,\bar{H}^2)$ where $\mathcal{L}(X)$ stands for the distribution of X ,

$$\bar{H} = (h_{ij}) \ , \ h_{ij} = \tfrac{1}{2}\left.\frac{\partial h}{\partial\theta_{ij}}\right|_{\theta=I} \ , \text{ for } i \neq j \ , \ h_{ii} = \left.\frac{\partial h}{\partial\theta_{ii}}\right|_{\theta=I}$$

Now observe that $\bar{H}$ can be rewritten as

(4.9) $\qquad \bar{H} = \tfrac{1}{2}H^* + \tfrac{1}{2}\,\mathrm{Diag}\,(\bar{H}^*)$

where $H^* = \left.\left(\frac{\partial h}{\partial\theta_{ij}}\right)\right|_{\theta=I}$ and Diag (A) is the diagonal matrix with only the diagonal elements of $A = (0_{ij})$.

Hence we can rephrase Lemma 4.1 as

Theorem 4.2.

By the same conditions as Lemma 4.1 we have

(4.10) $\qquad \mathcal{L}\{\sqrt{2n}\,[h(\theta) - h(I)]\} \to N(\cdot;0, \Lambda^*)$

where $\Lambda^* = \mathrm{tr}\,\bar{H}^{*2} + 3\,\mathrm{tr}[\mathrm{Diag}\,(\bar{H}^*)]^2$.

Now the testing statistic $\tilde{\lambda}$, as given in (4.9), is a function of S which is distributed as $W(\cdot;\Sigma, N-r)$. Let $\theta = (N-r)^{-1}\,\Sigma^{-\frac{1}{2}}\,S\Sigma^{-\frac{1}{2}}$, then

(4.11) $\qquad h(\theta) = \log\tilde{\lambda}(\theta) = \log|(X'\Sigma^{-\frac{1}{2}}\theta^{-1}\Sigma^{-\frac{1}{2}}X)^{-1}| - \log|B\Sigma^{\frac{1}{2}}\theta\Sigma^{\frac{1}{2}}B'|$.

It is obvious that $h(\theta)$ is scale invariant and the conditions for Theorem 4.2 are satisfied. Hence for the asymptotic non-null distribution of $\tilde{\lambda}$ we need only find $h(I)$ and Λ^* . It is easy to see that

(4.12) $$h(I) = \log\tilde{\lambda}(\theta = I) = \log \frac{|(X'\Sigma^{-1}X)^{-1}|}{|B\Sigma B'|} .$$

We next evaluate Λ^* . From (4.11) we have

$$\frac{\partial h(\theta)}{\partial h_{ij}} = - \frac{\partial}{\partial\theta_{ij}} \log|X'\Sigma^{-\frac{1}{2}}\theta^{-1}\Sigma^{-\frac{1}{2}}X| - \frac{\partial}{\partial\theta_{ij}} \log|B\Sigma^{\frac{1}{2}}\theta\Sigma^{\frac{1}{2}}B'|$$

$$= d_1 + d_2 \text{ , say .}$$

It can be shown that

$$d_1|_{\theta=I} = (\Sigma^{-\frac{1}{2}}X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-\frac{1}{2}})_{ji} ,$$

and

$$d_2|_{\theta=I} = -(\Sigma^{\frac{1}{2}}B'(B\Sigma B')^{-1}B\Sigma^{\frac{1}{2}})_{ji} .$$

Hence we have

(4.13) $$\bar{H}^* = \left(\frac{\partial h}{\partial\theta_{ij}}\right)_{\theta = I} = \Sigma^{-\frac{1}{2}}X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-\frac{1}{2}} - \Sigma^{\frac{1}{2}}B'(B\Sigma B')^{-1}B\Sigma^{\frac{1}{2}} .$$

From the previous results we conclude that under H_1

$$\mathcal{L}\{\sqrt{2n}\,[h(\theta) - h(I)]\} \to N(\cdot;0,\Lambda^*)$$

where

$$\Lambda^* = 2[m - \mathrm{tr}(X'\Sigma^{-1}X)^{-1}(B\Sigma B')^{-1}] + \mathrm{tr}[\mathrm{Diag}\,(\bar{H}^*)]^2,$$

$h(\theta)$, $h(I)$ were given in (4.11), (4.12) respectively, and $\bar{H}^*$ given in (4.13) .

Further as a sidelight of the S. S. model we also show the following:

Theorem 4.3.

The simple structure covariance $\Sigma = X\Gamma X' + Z\theta Z'$ will include the uniform covariance $\Sigma = \sigma^2[(1-\rho)I + \rho ee']$ as a special case with

$$\Gamma = \sigma^2 B[(1-\rho)I + \rho ee']B'$$

$$\theta = \sigma^2 D[(1-\rho)I + \rho ee']D'$$

if and only if

$$X'ee'Z = 0$$

where I is an identity matrix and e a $p\times 1$ vector consisting of all 1's .

<u>Proof:</u>

With Γ, θ specified as above, we have

$$\Sigma = \sigma^2\{XB[(1-\rho) + \rho ee'] B'X' + ZD[(1-\rho)I + \rho ee'] D'Z'\} .$$

By virtue of the fact that $X(X'X)^{-1}X' + Z(Z'Z)^{-1}Z' = I$ and $X'ee'Z = 0$ we have $\Sigma = \sigma^2 [(1-\rho)I + \rho ee']$.

Conversely, suppose $X\Gamma X' + Z\theta Z' = \sigma^2 [(1-\rho)I + \rho ee']$. Since $(X,Z)^{-1} = \binom{B}{D}$, we have $\begin{pmatrix}\Gamma & 0\\ 0 & \theta\end{pmatrix} = \binom{B}{D}\sigma^2 [(1-\rho)I + \rho ee'] (B',D')$ which implies $X'ee'Z = 0$. Hence the result.

It is to be noted that the condition $X'ee'Z = 0$ is equivalent to the requirement that e is in the space generated by X or by Z.

5. <u>Parameter Estimates and the Predictive Covariance (S. S.)</u>

The MLE's of τ, Γ and θ were given in (4.2). We now consider Bayesian estimates. With a convenient prior

$$g(\tau,\Gamma^{-1},\theta^{-1}) \propto |\Gamma|^{(m+1)/2} |\theta|^{(p-m+1)/2} ,$$

the joint posterior density of τ,Γ^{-1} and θ^{-1} is

$$P(\tau,\Gamma^{-1},\theta^{-1}) \propto |\Gamma^{-1}|^{(m+1)/2} |\theta^{-1}|^{(N-p+m-1)/2} \tag{5.1}$$

$$\cdot\exp\left(-\tfrac{1}{2}\mathrm{tr}\begin{bmatrix}\Gamma & 0\\ 0 & \theta\end{bmatrix}^{-1}\begin{bmatrix}BY-\tau A\\ DY\end{bmatrix}\begin{bmatrix}BY-\tau A\\ DY\end{bmatrix}'\right) .$$

From this density Geisser (1970) showed that a posteriori $\tau\sim D(\cdot;T_1,AA',BSB',N)$, and hence the posterior expectation of τ if it exists is T_1 . It is clear that Γ^{-1} and θ^{-1} are, a posteriori, independently distributed as $W(\cdot;(BSB')^{-1}, N-r)$ and $W(\cdot;(DYY'D')^{-1}, N)$ respectively. Thus by applying Lemma 1.2 we have

$$(5.2)\qquad \begin{cases} E(\Gamma) = (N-r-m-1)^{-1}\ BSB' \\ E(\theta) = (N-p+m-1)^{-1}\ DYY'D' \end{cases}$$

We next consider the covariance of the predictive distribution. Geisser (1970) showed that with $V_1 = BV$, $V_2 = DV$, the predictive density of V is $f(V) = f(V_1)\ f(V_2)$ where

$$(5.3)\qquad \begin{cases} V_1 \sim D(\cdot;T_1F,M,\ BSB',\ N+K-r) \\ V_2 \sim D(\cdot;0,\ I,\ DYY'D',\ N+K) \quad . \end{cases}$$

From Theorem 1.1 and $\binom{B}{D}^{-1} = (X,Z)$ we have

$$(5.4)\qquad Cov(V) = (N-m-r-1)^{-1}\ XBSB'X' \otimes M^{-1} + (N-p+m-1)^{-1}\ ZDYY'D'Z' \otimes I_K \quad .$$

We note that $M^{-1} = I + F'(AA')^{-1}F$ and

$$(5.5)\qquad [Cov(V)]^{-1} = (N-m-r-1)\ X(X'SX)^{-1}\ X' \otimes M + (N-p+m-1)Z(Z'YY'Z)^{-1}Z' \otimes I_K \ .$$

6. <u>Marginal and Conditional Predictive Densities (S.S.)</u>

In this section we shall consider predictive inference for the case $K = 1$. From (5.3) we have the predictive density of V

$$(6.1)\qquad f(V) \propto [1 + V_2'\ \ (DYY'D')^{-1}V_2]^{-(N+1)/2} \cdot [1 + M(V_1-T_1F)'(BSB')^{-1}(V_1-T_1F)]^{-(N+1-r)/2} \quad .$$

The density above can also be obtained by considering V_1 and V_2 to be independent and normally distributed conditional on u and w i.e.

$$F(V_1,V_2|u,w) = N(V_1;T_1F,\ (uM)^{-1}\ BSB')\cdot N(V_2;0,\ w^{-1}DYY'D')$$

where $F(\cdot|\cdot)$ denotes the distribution function and u, w are independently distributed as $\chi^2(N+1-r-m)$ and $\chi^2(N+1-p+m)$ respectively. Integration with respect to u and w yields (6.1).

Thus forming the joint density of V_1,V_2, u and w and then transforming from $\tilde{V} = \binom{B}{D}V$, u,w, to V,u,w we have $f(V,u,w)=f(V|u,w)\ g(u,w)$

where $F(V|u,w) = N(\cdot;XT_1F,\Lambda_s)$ for $\Lambda_s=(uM)^{-1}XBSB'X'+w^{-1}ZDYY'D'Z'$ and $g(u,w)$ is the product of two independent χ^2 densities.

Making the transformation $\frac{(N+1-p+m)u}{(N+1-r-m)w} = t, w = z$ and integrating out z we have

$$(6.2)\qquad f(V,t) = C_0\ \gamma^{(N+1-r)/2}\ t^{(N+1-r)/2-1}$$
$$\cdot\Gamma|\tfrac{1}{2}(2N+2-r)|\ |1+\gamma t+(V-XT_1F)'J(V-XT_1F)|^{-(2N+2-r)/2}$$

where

$$(6.3)\qquad \begin{cases} C_0=\pi^{-p/2}\ M^{m/2}|BSB'|^{-\frac{1}{2}}|DYY'D'|^{-\frac{1}{2}}\Gamma^{-1}|\tfrac{1}{2}(N+1-r-m)|\ \Gamma^{-1}|\tfrac{1}{2}(N+1-p+m)|\cdot\mathrm{mod}\left|\begin{pmatrix} B\\ D\end{pmatrix}\right| \\ \gamma = (N+1-p+m)^{-1}\ (N+1-r-m) \\ J = t\gamma MX(X'SX)^{-1}X' + Z(Z'YY'Z)^{-1}Z' \end{cases}$$

Thus an alternative representation for the predictive density is

$$(6.4)\qquad f(V) = \int_0^\infty f(V|t)\ g(t)\ dt$$

where

$$(6.5)\qquad \begin{cases} F(V|t) = T(\cdot;XT_1F,\ (1+\gamma t)\ (2N+2-r-p)^{-1}\ J^{-1},\ 2N+2-r) \\ t \sim F(\cdot;N+1-r-m,\ N+1-p+m)\ , \end{cases}$$

so that the predictive density of V is expressed as an average of multivariate T density over an F-density.

We next consider the marginal density of $V^{(1)}$. Let $J = \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22}\end{pmatrix}$ and V, X, Z be partitioned as before. From Lemma 3.1 we have

$$(6.6)\qquad f(V^{(1)}) = \int_0^\infty f(V^{(1)}|t)\ g(t)dt$$

where

$$(6.7)\qquad \begin{cases} F(V^{(1)}|t) = T(\cdot;X^{(1)}T_1F,\ (1+\gamma t)(2N+2-r-p)^{-1}\ J_{11.2}^{-1}\ ,\ 2N+2-r-p_2)\ , \\ t \sim F(\cdot;N-1-r-m,\ N+1-p+m) \\ J_{11.2} = J_{11} - J_{12}\ J_{22}^{-1}\ J_{21}\ . \end{cases}$$

Hence the marginal density of $V^{(1)}$ is of the same form as that of V and with expectation $X^{(1)}\ T_1F$. By interchanging the superscript

1 and 2 we obtain the marginal density of $V^{(2)}$.

We next consider conditional density of $V^{(2)}$ given $V^{(1)}$. From (6.2) and Lemma 3.1 we have

$$(6.8)\qquad f(V^{(2)}|V^{(1)}) = \int_0^\infty f(V^{(2)}|V^{(1)}, t)\, g(t|V^{(1)})\, dt$$

where

$$(6.9)\qquad \begin{cases} F(V^{(2)}|V^{(1)}, t) = T(\cdot; \mu_{s2.1}, b(2N+2-r-p_2)^{-1}, 2N+2-r) \\ \mu_{s2.1} = X^{(2)}T_1F - J_{22}^{-1} J_{21} (V^{(1)}-X^{(1)}T_1F) \\ b = 1 + \gamma t + (V^{(1)}-X^{(1)}T_1F)' J_{11.2}(V^{(1)}-X^{(1)}T_1F) \end{cases}$$

and

$$(6.10)\qquad g(t|V^{(1)}) = [f(V^{(1)})]^{-1}\pi^{p_2/2}\, t^{(N+1-r)/2-1}\,\Gamma[\tfrac{1}{2}(2N+2-r-p_2)]\, b^{-(2N+2-r-p_2)/2} \cdot C_o |J_{22}|^{-\frac{1}{2}}$$

From (6.8) we also obtain the conditional expectation of $V^{(2)}$ given $V^{(1)}$

$$(6.11)\qquad E(V^{(2)}|V^{(1)}) = \int_0^\infty \mu_{s2.1}\, g(t|V^{(1)})\, dt \quad .$$

We hence obtain a one-dimensional integral representation for the conditional density of $V^{(2)}$ given $V^{(1)}$ and the conditional expectation as given in (6.8) and (6.11) respectively.

By (1.19) and Theorem 1.1 we obtain the covariance of $V^{(2)}$ given $V^{(1)}$ as

$$(6.12)\qquad \begin{aligned} Cov(V^{(2)}|V^{(1)}) = {} & E_{t|V^{(1)}} (2N+1-r-p_2)^{-1}\, bJ_{22}^{-1} \\ & + Cov_{t|V^{(1)}}\, J_{22}^{-1}J_{21}\, (V^{(1)}-X^{(1)}T_1F) \end{aligned}$$

which is essentially a one-dimensional integral.

Some computational considerations: In computing (6.11) we note that

$$(6.13)\qquad g(t|\ddot{V}^{(1)}) = \frac{t^{(N+1-r)/2^{-1}}\, b^{-(2N+2-r-p_2)/2}\, |J_{22}|^{-\frac{1}{2}}}{\int_0^\infty t^{(N+1-r)/2-1}\, b^{-(2N+2-r-p_2)/2} |J_{22}|^{-\frac{1}{2}}\, dt} \, .$$

Also, by the spectral representation theorem we have

$$(6.14)\quad \begin{cases} J_{22} = \Phi_1' D(1+|\gamma Mt-1|\lambda_i)\psi_1' [X^{(2)}(X'SX)^{-1}X^{(2)'}+Z^{(2)}(Z'YY'Z)^{-1}Z^{(2)'}] \\ (V^{(1)}-X^{(1)}T_1F)'J_{11.2}(V^{(1)}-X^{(1)}T_1F)=(V^{(1)}-X^{(1)}T_1F)'\Phi_2 D^{-1}(1+\left[\frac{1}{\gamma Mt}-1\right]\ell_i)\psi_2 \\ \quad \cdot[X^{(1)}BSB'X^{(1)'} + Z^{(1)}DYY'D'Z^{(1)'}]^{-1}(V^{(1)}-X^{(1)}T_1F) \end{cases}$$

where $D(a_i)$ is a diagonal matrix with elements a_i , λ_i are characteristic roots of $X^{(2)}(X'SX)^{-1}X^{(2)'}[X^{(2)}(X'SX)^{-1}X^{(2)'}+Z^{(2)}(Z'YY'Z)^{-1}Z^{(2)'}]^{-1}$, $\Phi_1\psi_1 = I = \Phi_2\psi_2$ and Φ_1,ψ_1 and Φ_2,ψ_2 are the orthogonal matrices reducing the matrices to their canonical forms. Hence we need not evaluate determinants and invert matrices for each value of t .

For the calculation of the covariance of $V^{(2)}$ given $V^{(1)}$ as given in (6.12), a similar utilization of the spectral representation theorem will also preclude the computation of determinants and inverses for each value of t .

7. Approximations (S.S.)

Since conditional on an F variate the predictive density of V is a multivariate T as given in (6.5), we may approximate $F(V)$ by

$$(7.1)\qquad F(V|\hat{t}) = T(\cdot;XT_1F,(1 + \gamma\hat{t})(2N+2-r-p)^{-1}J^{-1}(\hat{t}), 2N+2-r)$$

where

$$\hat{t} = \frac{N+1-p+m}{N-1-p+m} \quad \text{i.e., the mean of the F-distribution ; or}$$

$$\hat{t} = \frac{N-r-m-1}{N+1-r-m}\,\frac{N+1-p+m}{N+3-p+m} \quad \text{i.e., the mode of the F-distribution,}$$

and $J(\hat{t})$ is the value of J evaluated at $t = \hat{t}$.

From (7.1) we have

$$(7.2)\qquad F(V^{(2)}|V^{(1)}) \simeq T(\cdot;\mu_{s2.1}(\hat{t}), b(\ddot{t})(2N+-r-p_2)^{-1}J_{22}^{-1}(\hat{t}),2N+2-r) .$$

Thus an approximate point estimate of $V^{(2)}$ given $V^{(1)}$ is $\mu_{s2.1}(\hat{t})$ and an approximate predictive region can be obtained through

(7.3) $$\left(V^{(2)}-\mu_{s2.1}(\hat{t})\right)'\left[b^{-1}(\hat{t})J_{22}(\hat{t})\right]\left(V^{(2)}-\mu_{s2.1}(\hat{t})\right)$$

being approximately distributed as

$$\frac{p_2}{2N+2-r-p_2} F(p_2, 2N+2-r-p_2) \quad .$$

Similar arguments may be employed for marginal inference on $V^{(i)}$, for $i = 1, 2$, from the approximation given in (7.1).

An alternative approximation for $f(V^{(2)}|V^{(1)})$ is

(7.4) $$f(V^{(2)}|V^{(1)}) \approx T(\cdot;\mu_{s2.1}(t_o), b(t_o)(2N+2-r-p_2)^{-1}J_{22}^{-1}(t_o), 2N+2-r)$$

where t_o is the value of t which maximizes $g(t|V^{(1)})$ or $g(\log t|V^{(1)})$.

In the rest of this section we assume that Γ and θ are essentially known, i.e., replaced by estimates, unbiased estimates MLE's or posterior expectations.

The density of V conditional on τ , Γ^{-1} and θ^{-1} is

(7.5) $$f(V|\tau,\Gamma^{-1},\theta^{-1}) \propto |\Gamma^{-1}|^{K/2}\ |\theta^{-1}|^{K/2} \exp\left[-\tfrac{1}{2}\mathrm{tr}\begin{pmatrix}\Gamma & 0\\ 0 & \theta\end{pmatrix}^{-1}\begin{pmatrix}BV-\tau F\\ 0\end{pmatrix}\begin{pmatrix}BV-\tau F\\ 0\end{pmatrix}'\right] \quad .$$

Combining (5.1) and (7.5), with Γ^{-1}, θ^{-1} assumed known, and integrating out τ we have

(7.6) $$f(V|\Gamma^{-1},\theta^{-1}) \propto \exp\left[-\tfrac{1}{2}\mathrm{tr}\theta^{-1}D(YY'+VV')D'-\tfrac{1}{2}\mathrm{tr}\Gamma^{-1}(BV-T_1F)M(BV-T_1F)'\right]$$

By the transformation $\begin{pmatrix}B\\ D\end{pmatrix} V = \begin{pmatrix}V_1\\ V_2\end{pmatrix}$ and a property of normal distributions we have

(7.7) $$F(V|\Gamma^{-1},\theta^{-1}) = N(\cdot; X\Gamma X'\otimes M^{-1} + Z\theta Z'\otimes I_K) \ .$$

Thus we see that with Γ, θ replaced by their estimates inference on V as well as $V^{(2)}$ given $V^{(1)}$ can be obtained through standard normal theory.

8. Predictive Modes (S.S.)

It was shown by Geisser (1970) that XT_1F is the mode of the predictive distribution of V. In this section we present a numerical procedure for calculating the mode of $V^{(2)}$ given $V^{(1)}$ when $K = 1$.

The predictive density of V when $K = 1$ can be written as

$$f(V) \propto [1 + (V-XT_1F)'C(V-XT_1F)^{-(N+1)/2}[1 + (V-XT_1F)'E(V-XT_1F)]^{-(N+1-r)/2} \tag{8.1}$$

where

$$\begin{cases} C = Z(Z'YY'Z)^{-1}Z' = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \\ E = MX(X'SX)^{-1} X' = \begin{pmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix} \end{cases} \tag{8.2}$$

By Lemma 3.1 we have conditional density of $V^{(2)}$ given $V^{(1)}$

$$\begin{aligned} f(V^{(2)}|V^{(1)}) \propto\ & [b_1+(V^{(2)}-\mu_{.1})'C_{22}(V^{(2)}-\mu_{.1})]^{-(N+1)/2} \\ & \cdot[b_2 + (V^{(2)}-\mu_{.2})'E_{22}(V^{(2)}-\mu_{.2})]^{-(N+1-r)/2} \end{aligned} \tag{8.3}$$

where

$$\begin{cases} b_1 = 1 + (V^{(1)}-X^{(1)}T_1F)'C_{11.2}(V^{(1)}-X^{(1)}T_1F) \\ \mu_{.1} = X^{(2)}T_1F - C_{22}^{-} C_{21}(V^{(1)}-X^{(1)}T_1F) \\ C_{11.2} = C_{11} - C_{12} C_{22}^{-} C_{21} \\ \mu_{.2} = X^{(2)}T_1F - E_{22}^{-} E_{21}(V^{(1)}-X^{(1)}T_1F) \\ b_2 = 1 + (V^{(1)}-X^{(1)}T_1F)' E_{11.2} (V^{(1)}-X^{(1)}T_1F) \\ E_{11.2} = E_{11}-E_{12} E_{22}^{-} E_{21} \end{cases}$$

with the convention that $A_{22}^{-} = 0$ if $A_{22} = 0$.

Using the same arguments as given in section 3 we see that the mode of conditional distribution of $V^{(2)}$ given $V^{(1)}$ will satisfy

$$V_m^{(2)} = (E_{22} + \lambda C_{22})^{-1} (E_{22}\mu_{.2} + \lambda C_{22}\mu_{.1}) . \tag{8.5}$$

The existence of the inverse of $E_{22} + C_{22}$ is assured by the fact that $E + C$ is p.d. Hence we can search for a λ^* such that $V_m^{(2)}(\lambda^*)$ maximizes $f(V^{(2)}|V^{(1)})$. The mode of $V^{(2)}$ given $V^{(1)}$ is thus obtained in a one-dimensional search over λ .

Instead of substituting $V_m^{(2)}$ into (8.3) and search for $V^{(2)}$ which maximizes $f(V^{(2)}|V^{(1)})$, we consider maximizing $h(\lambda)=f(V_m^{(2)}|V^{(1)})$, or equivalently $\log h(\lambda)$. Further $\frac{\partial}{\partial\lambda} \log h(\lambda) = 0$ yields

$$\lambda = \frac{Q_N}{Q_D} = \ell(\lambda) \tag{8.6}$$

where

$$\left\{\begin{aligned} Q_N &= -\frac{N+1}{2} b_1\alpha_1 - \frac{N+1-r}{2} b_2\alpha_2 \\ \alpha_1 &= (\mu_{\cdot 1}-\mu_{\cdot 2})'E_{22}\big[(E_{22}+\lambda C_{22})^{-2} C_{22}^2(\lambda^{-1}E_{22} + C_{22})^{-1} \\ &\quad + (\lambda^{-1}E_{22} + C_{22})^{-1} C_{22}(E_{22} + \lambda C_{22})^{-2}C_{22}\big] E_{22}(\mu_{\cdot 1}-\mu_{\cdot 2}) \\ \alpha_2 &= (\mu_{\cdot 1}-\mu_{\cdot 2})'C_{22}\big[(E_{22} + \lambda C_{22})^{-2} E_{22}C_{22}(\lambda^{-1}E_{22} + C_{22})^{-1} \\ &\quad + (\lambda^{-1}E_{22} + C_{22})^{-1} E_{22}(E_{22} + \lambda C_{22})^{-2}C_{22}\big]E_{22}(\mu_{\cdot 1}-\mu_{\cdot 2}) \\ Q_D &= (V_m^{(2)} - \mu_{\cdot 1})'C_{22}(V_m^{(2)}-\mu_{\cdot 1})\alpha_1\lambda^{-1} + (V_m^{(2)}-\mu_{\cdot 2})'E_{22}(V_m^{(2)}-\mu_{\cdot 2})\alpha_2\lambda^{-1} . \end{aligned}\right. \tag{8.7}$$

The relation (8.6) will enable us to find recursively all the stationary points of $h(\lambda)$. After obtaining all the stationary points of $h(\lambda)$ we then substitute $V_m^{(2)}(\lambda)$ into (8.3) to find the mode.

We next consider the special case when $p_2 = 1$, i.e. we have only one point to predict. Now (8.3) becomes

$$f(V^{(2)}|V^{(1)}) \propto \left[1 + C^*_{22} (V^{(2)}-\mu_{\cdot 1})^2\right]^{-(N+1)/2} \cdot\left[1 + E^*_{22}(V^{(2)}-\mu_{\cdot 2})^2\right]^{-(N+1-r)/2} \tag{8.8}$$

where

$$\left\{\begin{aligned} C^*_{22} &= b_1^{-1} C_{22} \\ E^*_{22} &= b_2^{-1} E_{22} \end{aligned}\right. . \tag{8.9}$$

Then $\frac{\partial}{\partial V^{(2)}} \log f(V^{(2)}|V^{(1)}) = 0$ implies

$$V^{(2)^3} + \beta V^{(2)^2} + \epsilon V^{(2)} + \delta = 0 \tag{8.10}$$

where

$$(8.11)\quad \begin{cases} \beta = -\alpha^{-1}E^*_{22}C^*_{22}\,[(3N+1)\ \mu_{.1} + (3N+2)\ \mu_{.2}] \\ \alpha = (2N+1)\ E^*_{22}C^*_{22} \\ \epsilon = \alpha^{-1}E^*_{22}C^*_{22}\,[N\mu^2_{.1} + (N+1)\mu^2_{.2} + (4N+2)\mu_{.1}\mu_{.2}] + (N+1)C^*_{22}+NE^*_{22} \\ \delta = -\alpha^{-1}E^*_{22}C^*_{22}\mu_{.1}\mu_{.2}[N\mu_{.1} + (N+1)\mu_{.2}] - N\mu_{.2}E^*_{22}-(N+1)\mu_{.1}C^*_{22}\ . \end{cases}$$

Let $p^* = \epsilon - \frac{\beta^2}{3}$, $q^* = \delta - \frac{\beta\epsilon}{3} + \frac{2\beta^3}{27}$.

(A) If $\Delta = 4p^{*3} + 27q^{*2} > 0$, the mode is

$$V_m^{(2)} = (A^*)^{1/3} + (B^*)^{1/3}$$

where $A^* = -\frac{q^*}{2} + \left(\frac{q^{*2}}{4} + \frac{p^{*3}}{27}\right)^{\frac{1}{2}}$, $B^* = -\frac{q^*}{2} - \left(\frac{q^{*2}}{4} + \frac{p^{*3}}{27}\right)^{\frac{1}{2}}$.

(B) If $\Delta = 0$, then the three roots of (8.10) are $2\left(-\frac{q^*}{2}\right)^{\frac{1}{3}}$, $\left(\frac{q^*}{2}\right)^{\frac{1}{3}}$, $\left(\frac{q^*}{2}\right)^{\frac{1}{3}}$ respectively. Then

$$(8.12)\quad \begin{cases} \text{(a)} \quad \text{if } q^* > 0, \text{ the mode is } \left(\frac{q^*}{2}\right)^{\frac{1}{3}} , \\ \text{(b)} \quad \text{if } q^* < 0, \text{ the mode is } 2\left(-\frac{q^*}{2}\right)^{\frac{1}{3}} , \\ \text{(c)} \quad \text{if } q^* = 0, \text{ the mode is } 0\ . \end{cases}$$

(C) If $\Delta < 0$, then the three roots of (8.10) are $2\left(-\frac{p^*}{3}\right)^{\frac{1}{2}}$ Cos $\varphi/3$, $2\left(-\frac{p^*}{3}\right)^{\frac{1}{2}}$ Cos $(60^\circ - \varphi/3)$ and $-2\left(\frac{p^*}{3}\right)^{\frac{1}{2}}$ Cos$(60^\circ + \varphi/3)$ respectively where φ is determined by Cos $\varphi = \frac{q^*(27)^{\frac{1}{2}}}{2p^*(-p^*)^{\frac{1}{2}}}$, or tan $\varphi = \frac{-(-\Delta)^{\frac{1}{2}}}{q^*(27)^{\frac{1}{2}}}$ on the condition that φ is taken in the first or second quadrant according as $q^* < 0$ or > 0 . Since this is a bimodal case, we need only substitute two extreme roots into (8.8) to find the global mode. We note in passing that the necessary and sufficient condition for (8.8) to be bimodal is $\Delta < 0$.

We next consider the joint mode of the distribution of $V^{(2)}$ and t given $V^{(1)}$. From (6.9) and (6.10) it is easily seen that the joint mode is $(\mu_{s2.1}(t^*), t^*)$ where t^* maximizes

$$g(t) = t^{(N+1-r)/2\ -1}\ b^{-(2N+2-r)/2} \quad . \tag{8.13}$$

The marginal components of the above joint mode corresponding to $V^{(2)}$, namely $\mu_{s2\cdot1}(t^*)$, can also be used as an estimator of $V^{(2)}$ given $V^{(1)}$ in much the same way as the "generalized maximum likelihood"estimator of DeGroot (1970 p. 236).

Since in (8.14) $b > 1$ and the negative power of b is twice as big as the positive power of t , the maximum of g(t) should occur in a small interval $(0,1]$ or $(0,2]$. Thus in practice we may just plot g(t) for $t\epsilon(0,2]$ and find t^* . Hence $\mu_{s2.1}(t^*)$ will appear to be, in practice, an appropriate point estimate for $V^{(2)}$ given $V^{(1)}$.

Similar to the search for the mode we can also let $\frac{\partial}{\partial t} \log g(t) = 0$ which yields the relation

$$t = R(t) \tag{8.14}$$

where

$$\begin{cases} R(t) = b(N-2)(2N+1)\left[\gamma+(V^{(1)}-X^{(1)}T_1F)'\dot{J}_{11.2}(V^{(1)}-X^{(1)}T_1F)\right]^{-1} \\ \dot{J}_{11.2} = \dot{J}_{11} - \dot{J}_{12}\ J_{22}^{-1}\ J_{21} - J_{21}\ \dot{J}_{22}^{-1}\ J_{21}-J_{12}J_{22}^{-1}\ \dot{J}_{21} \\ \dot{J}_{ij} = \frac{\partial}{\partial t}\ J_{ij} \end{cases} \tag{8.15}$$

Further (8.14) may then be solved recursively for the stationary points.

9. Final Remarks.

Although some of these procedures appear to be algebraically complicated, these mean and modal values are often not at all difficult to calculate by computer. In a subsequent paper we shall give examples of these calculations and also compare the methods of partial prediction generated here with a class of much simpler heurestic procedures, one of which we term quasi-least square methods, on several sets of data.

References

(1) Anderson, T. W. (1958): An Introduction to Multivariate Statistical Analysis. John Wiley and Sons, Inc., New York.

(2) Cramer, H. (1964): Mathematical Method of Statistics. Princeton University Press, Princeton, N.J.

(3) DeGroot, M. H. (1970): Optimal Statistical Decisions. McGraw-Hill, Inc.

(4) Dickey, James, M. (1967): Matricvariate Generalizations of the Multivariate t Distribution and the Inverted Multivariate t Distribution. Ann. Math. Stat., 38, pp. 511-518.

(5) Geisser, S. (1966): Predictive Discrimination. Multivariate Analysis Proceedings of an International Symposium held in Dayton, Ohio, June 14-19, 1965. Academic Press, pp. 149-163.

(6) Geisser, S. (1970): Bayesian Analysis of Growth Curves Sankhya, Series A, 32, pp. 53-64.

(7) Geisser, S. and Cornfield, J. (1963): Posterior Distributions for Multivariate Normal Parameters. J. Roy. Stat. Soc., Series B, 25, pp. 368-376.

(8) Khatri, C. G. (1966): A note on MANOVA Model Applied to Problems in Growth Curve. Ann. Inst. Stat. Math., 18, pp. 75-86.

(9) Lachenbruch, P. A. (1968): On Expected Probabilities of Misclassification in Discriminant Analysis, Necessary Sample Size, and a Relation with the Multiple Correlation Coefficient. Biometrics, 24, pp. 823-833.

(10) Lindley, D. V. (1970): A Bayesian Solution for Some Educational Prediction Problems, III. Educational Testing Service Research Bulletin, RB - 70 -33.

(11) Olkin, I. and Press, S. J. (1969): Testing and Estimation for a Circular Stationary Model. Ann. Math. Stat., 40, pp. 1358-1373.

(12) Rao, C. R. (1965): The Theory of Least Squares when the Parameters are Stochastic and its Application to the Analysis of Growth Curves. Biometrika, 52, pp. 447-458.

(13) Rao, C. R. (1967): Least Squares Theory using an Estimate Dispersion Matrix and its Application to Measurement of Signals. Proceedings of the Fifthe Berkeley Symposium on Math. Stat. and Probability, 1, pp. 355-372.

(14) Zelen, M. and Federer, W. T. (1965): Application of the Calculus for Factorial Arrangements III. Analysis of Factorials with Uneqaul Numbers of Observations. Sankhyā, Series A, 27, pp. 383-400.